

TheDataWeb & DataFerrett



Bill Hazard
The DataWeb Applications Office
U.S. Census Bureau

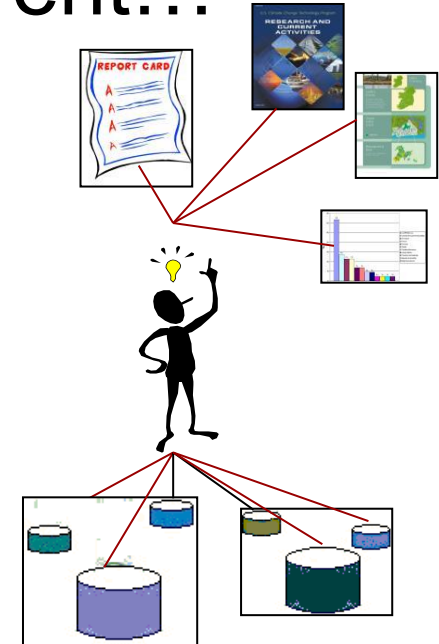
The Census Bureau's Breakthrough Technologies

- ✓ **TheDataWeb** is a network of public and private databases -- a virtual data warehouse on the Internet.
- ✓ **DataFerrett** is a web browser that searches TheDataWeb and allows analysts to recode variables, add calculations and create custom tables.
- ✓ **HotReports** use DataFerrett to retrieve data from TheDataWeb to produce interactive web pages with custom tables, maps, graphs and text.

TheDataWeb Defined

An infrastructure for linking different databases into a single virtual data warehouse that supports different...

- ✓ vendors
- ✓ file structures
- ✓ data types
- ✓ file security models



DataFerrett Defined

DataFerrett is the interface for TheDataWeb.
It allows a user to:

- ✓ Browse (search & discover) all of the datasets that are in TheDataWeb
- ✓ Select variables from multiple datasets and integrate data on-the-fly
- ✓ Produce sophisticated analyses using tables, graphs, and maps
- ✓ Create customized HotReports, or web “mash ups”

What DataFerrett Can Do

- ✓ Data manipulation and variable recoding
- ✓ Match/merge files for you (pre-merged)
- ✓ Frequencies, cross- and multi-dimension (nested) tabulations
- ✓ Spreadsheet formulas
- ✓ Mapping & graphing
- ✓ Publishing to HTML and PDF

Simple Cross-Tabulation


Ferrett Tabulation
File Edit Format View Options Help

GO Get Data

	C1	C2	C3	C4	C5	C6	C7
R1		Total RECODE2	Not Hispanic	Hispanic			
R2	Total RECODE1	56,600	42,751	13,849			
R3	White alone	26,109	20,013	6,096			
R4	Black or African American alone	17,455	17,317	138			
R5	AIAN alone	214	214	0			
R6	Asian alone	4,874	4,874	0			
R7	Native Hawaiian and Other Pacific Islander alone	0	0	0			
R8	Some other race alone	7,267	68	7,199			
R9	Two or more major race groups	681	265	416			
R10							
R11							
R12							
R13							
R14							
R15							
R16							
R17							
R18							
R19							
R20							
R21							
R22							
R23							
R24							
R25							
R26							
R27							
R28							

GCR Responsible for grandchildren
GEOG-101 FIPS State Code
RAC1P Race1 recode
HISP Hispanic recode
RECODE1 Race Recode
RECODE2 Hispanic - Not Hispanic

Universe: (GCR in (1)) AND ((ST
Weight used: PWGTP
DataSet(s) selected: 2005

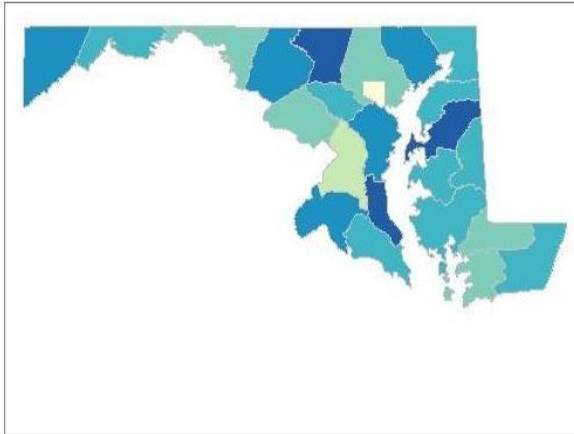


Data Visualization

Highlight spreadsheet rows or columns for:

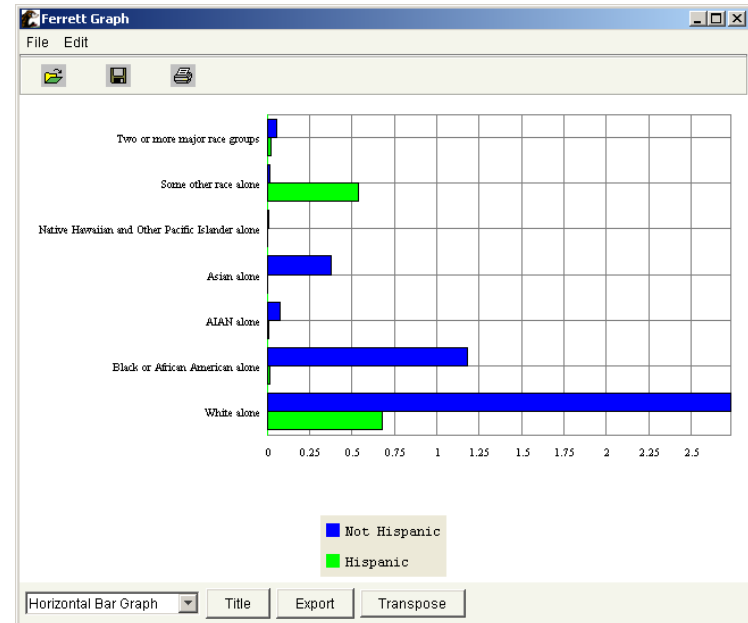
Percent Owner Occupied Housing Units: 2000

50+ -- 55
55+ -- 60
60+ -- 65
65+ -- 70
70+ -- 75
75+ -- 80
80+ -- 86



Mapping

Graphing



Powerful Tabulation Capabilities

Simple table layout for frequencies and trends;
supporting spreadsheet math for complex analysis

The screenshot shows the Ferrett Tabulation software window. The main table displays data for various education levels (rows R1-R25) categorized by labor force status (columns C2-C7). The table is structured as follows:

	C1	C2	C3	C4	C5	C6	C7
		Total In Labor Force		Employed		Unemployed	
		Male	Female	Male	Female	Male	Female
R1							
R2	Total - Education - In Labor Force	78,483,029	68,419,679	73,413,695	64,548,829	5,069,334	3,870,851
R3	Less Than 1st Grade	212,865	108,668	191,889	93,734	20,975	14,934
R4	1st,2nd,3rd Or 4th Grade	629,118	230,852	586,926	187,326	42,193	43,527
R5	5th Or 6th Grade	1,371,110	635,981	1,264,839	563,517	106,271	72,464
R6	7th Or 8th Grade	1,318,163	730,752	1,183,722	627,921	134,441	102,831
R7	9th Grade	1,701,630	1,119,725	1,491,006	919,928	210,624	199,797
R8	10th Grade	2,108,519	1,617,745	1,821,488	1,378,381	287,031	239,364
R9	11th Grade	2,844,985	2,136,417	2,454,361	1,834,758	390,624	301,659
R10	12th Grade No Diploma	1,161,697	770,492	1,045,690	703,612	116,007	66,880
R11	High School Grad-Diploma Or Equiv (GED)	23,969,853	20,081,948	22,099,431	18,812,676	1,870,422	1,269,271
R12	Some College But No Degree	14,563,011	14,279,620	13,653,180	13,553,140	909,831	726,479
R13	Associate Degree-Occupational/Vocational	3,371,443	3,608,261	3,173,004	3,455,721	198,439	152,540
R14	Associate Deg.-Academic Program	2,626,552	3,226,178	2,502,476	3,116,536	124,076	109,643
R15	Bachelor's Degree(ex:ba,ab,bs)	14,645,523	13,401,905	14,174,086	12,975,955	471,437	425,950
R16	Government - Federal	430,622	308,170	419,228	302,350	11,393	5,820
R17	Government - State	636,556	926,983	632,705	919,958	3,850	7,025
R18	Government - Local	1,093,046	1,857,165	1,080,539	1,820,737	12,507	36,428
R19	Private, For Profit	9,665,227	7,544,798	9,268,887	7,225,822	396,340	318,976
R20	Private, Nonprofit	673,979	1,669,439	660,287	1,634,133	13,692	35,306
R21	Self-Employed, Incorporated	1,037,620	356,345	1,024,559	352,166	13,062	4,179
R22	Self-Employed, Unincorporated	1,108,473	730,253	1,087,880	712,038	20,593	18,216
R23	Without Pay		0	8,752	0	8,752	0
R24	Master's Degree(ex:MA,MS,MEng,MEd,MSW)	5,056,012	5,094,156	4,920,866	4,983,933	135,146	110,223
R25	Professional School Deg(ex:MD,DDS,DVM)	1,570,478	782,566	1,545,646	762,301	24,832	20,264

- Massive Datasets
- Complex Nesting
- Formulas
- Hide columns or rows
- Use weighting variables

DataFerrett Users

Who should use DataFerrett?

- ✓ Intended for serious data users – Power Users
 - ✓ Users that want to “play” with the data
 - ✓ Users who cannot get what they need from the pre-defined data tables
 - ✓ Users needing quick, yet sophisticated tabulations

What We're Working On

- ✓ Calculating variances on-the-fly for microdata tabulations
- ✓ Advanced statistical methods – integrating the R statistical package
- ✓ Regressions
- ✓ Cluster analysis

Supports Multiple Data Types

(Needed for data integration)

Microdata: Individual transaction records, or survey response records. Data is often in multiple files (e.g. Household, Person, Diary, Geography files).

Aggregate (Macro) Data: Data that has already been tabulated. You must pick from variables or dimensions already tabulated (i.e. geography).

Timeseries: Data that can be tabulated by time periods from a cell in a table (poverty rate, or unemployment rate over time).

Longitudinal Data: follows people over time (microdata).

- Quick Facts
- American FactFinder
- Population Finder
- 2010 Census
- Economic Census
- Interactive Maps
- Training & Workshops
- Data Tools**
- Catalogs
- Publications

Stats in Action

Target uses statistical information from the American Community Survey (ACS) to understand more about communities and better serve their guests. The ACS provides detailed data on housing, housing values, people's incomes and more.

[See More](#)



Construction

U.S. Census Bureau

Data Access Tools

You are here: [Census.gov](#) > Data Access Tools

[Data Tools Main](#)

• Interactive Internet Tools

- [2010 Census Interactive Population Map](#) - Use this charts on your web site.
- [The American FactFinder](#) - This interactive application offers:
- [QuickFacts](#) - State and County QuickFacts provides
- [Censtats](#) - Applications available include: Census Tracts
- [Online Mapping Tools](#) - using TIGER and the American
- [US Gazetteer](#) - Place name, and ZIP code search
- [Population Dynamics Statistics](#) - This tool shows tab
- **[DataFerrett](#)** - A tool and data librarian that searches tabulation and business graphics. Current Population, Small Area Income Poverty Estimates, Population by Control data, and more. DataFerrett's newest tool, the [Community Economic](#) the Employment & Training Administration's Workfo



Contact Us

**Bill Hazard –
william.g.hazard@census.gov**

**DataFerrett Help Desk –
dsd.ferrett@census.gov**

DataFerrett Features

Merging Files

- DataFerrett automatically merges files for you.

Filtering records to restrict a universe

- In SAS you use a Data Step and an "If variable X = Y;" to keep variables for analysis.
- In DataFerrett you simply click on the values of the variables that you want to include.

Creating new variables

- In SAS or SPSS, you will create a variable in a Data Step or script with a statement like "If (variable X =Y) then variable Z = variable X + 22 etc.

DataFerrett actually supports data manipulation in 3 ways.

1. Creating/saving a variable as a simple recode. This collapses the values from a single variable to more useful classes.
2. Creating a new variable from several variables using "if then else" logic.
3. Creating new aggregate variables by using the spreadsheet(spreadsheet math) to manipulate rows and columns.

Saving programs/analysis

Variables and scripted transformations are saved in two ways:

1. Saving the data basket in a DataFerrett Session File (FSF)
2. Saving the layout of a table in a DataFerrett Tabulation File (FTF) *note: this also saves the FSF associated with the table.*